

海量实时广告流平台架构设计与实践

开讲日期：

授课讲师：

内容大纲

1. 广告系统介绍
2. 如何设计一个高性能的广告数据流处理平台
3. 广告数据管理平台架构设计

项目背景

在互联网上面，流量=变现。

流量变现的方式有很多种，不同的产品/平台，变现模式不一样。按大类分，无外乎三类：**卖广告、卖产品/服务、增值服务**。

今天重点说说广告的变现模式，并主要解决两个问题：

1. 广告从何来？
2. 怎样的广告策略才能让流量变现最大化？

广告来源可以有很多渠道，分两个大类：

1. 直投客户：来自自己销售，或代理公司。

这里的直投是指直接在流量方广告系统投放，可以自建也可以使用第三方广告系统。

在变现上，DAU千万级以上的产品就不用过于发愁了，不管是自建销售队伍还是走代理公司，一般都是广告主或代理公司争抢的TOP流量。只要想变现，还是不缺广告填充的，因为平台级产品是可以掰着手指头数的过来的。

而且有了这么大量级的用户，变现的玩法更多，如果哪家用户过千万级产品还苦于没钱赚，那负责变现的朋友自动下课好了。

2. DSP（或SSP、或联盟广告）

对于DAU规模在大几十万甚至百万级别的产品（垂直细分类产品除外，如金融类），流量有了一定的广告价值，可能会有广告主或代理公司主动找上门要求投放，但这类基本属于守株待兔型的，如果不主动出击，可能都解决不了温饱问题。

如果往前推个十年，网站只要稍微有点流量，不少都是靠在底部留个联系方式等广告上门，而且客户还真不少，因为那会的环境是狼少肉多。

特别对于流量规模更小的APP或个人网站，基本没有销售队伍或者销售能力，广告填充基本都是来自于DSP或联盟广告。

怎样的广告策略才能让流量变现最大化？

通过数据驱动，将数据作为生产资料，通过科学的方法，运用到业务经营过程中，并不断作出正向的反馈，促进业务优化提高。



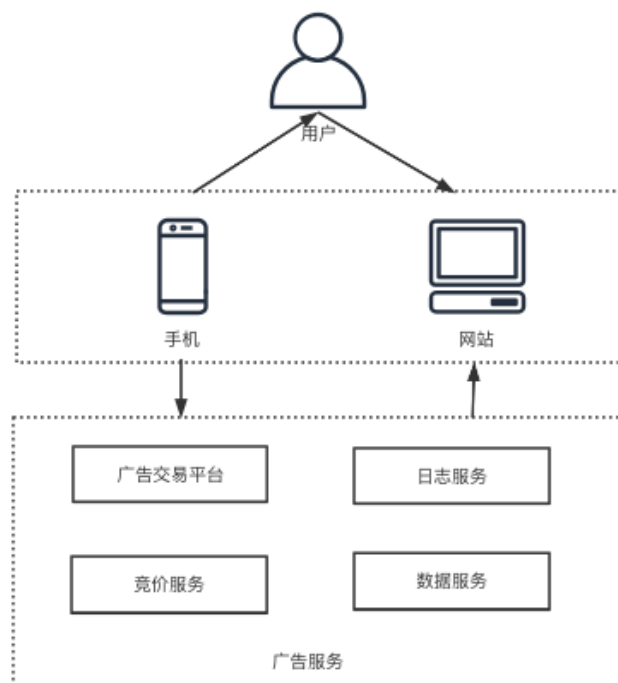
数据管理平台

数据管理平台是互联网广告体系中非常重要的一部分

- 是广告主实现自身数据增值的重要工具，广告主通过大数据分析消费者的行为，可以将广告更有针对性地传达给“刚需客户”。
- 通过数据统计分析用户的访问行为和喜好，将内容更有针对性地推送给用户，提高用户访问的粘性，增大网站的流量。
- 数据管理平台提供了多维度的查询接口，让广告主和运营人员能够实时的查看数据效果展示。

从数据角度来看，通过数据管理平台，需求方平台的广告投放的效果得到增强，并通过直观的数据展示，也让更多的广告主认可了广告投放平台，通过数据的应用为广告主和公司创造了收益。

DSP广告系统架构



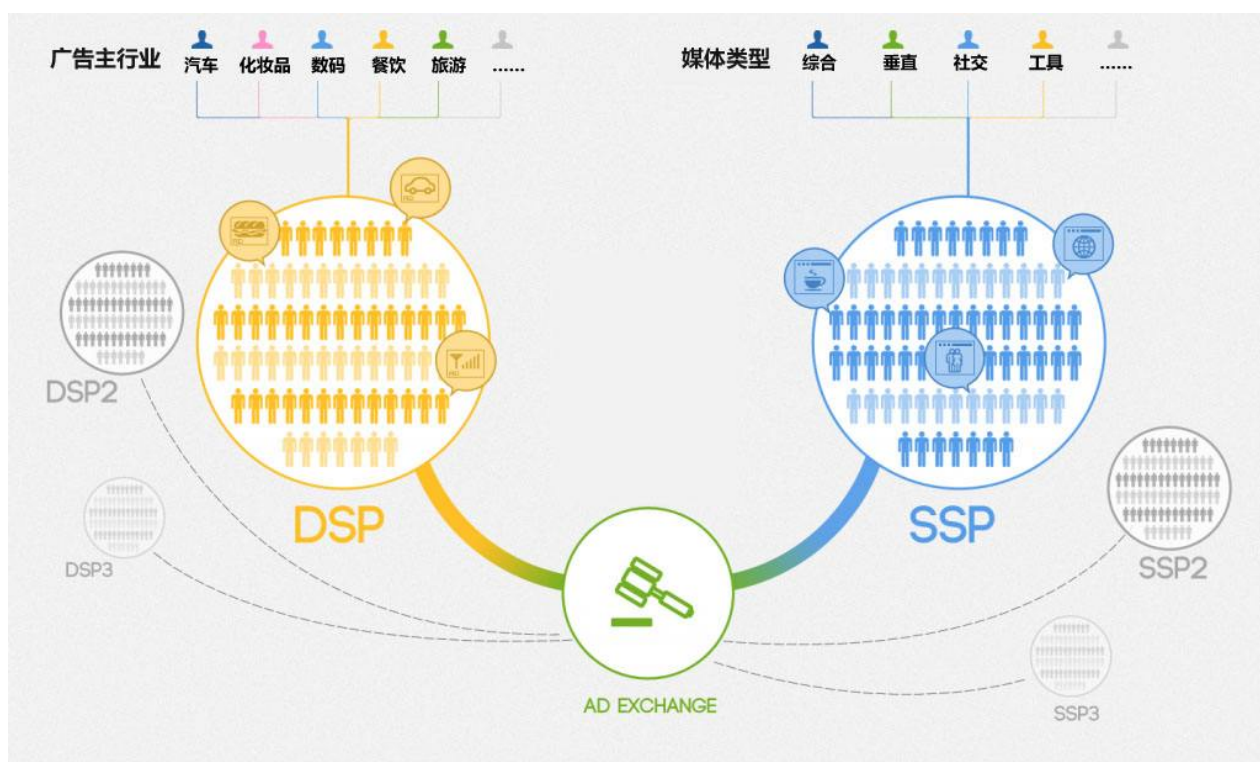
DSP是什么

DSP(Demand-Side Platform)，就是需求方平台，以精准营销为核心理念。允许广告客户和广告机构更方便地访问，以及更有效地购买广告库存。

DSP是干什么的

为需求方（即广告主或代理商）提供实时竞价投放平台，需求方可以在DSP平台上管理广告活动及其投放策略，包括目标受众的定向条件、预算、创意、出价等等，DSP可以通过技术和算法优化投放效果，提供数据报告，使需求方清晰明了的看到自己的广告效果。

说白了就是把产品的广告直接送到目标客户面前。



DSP的目标

使用程序的方式进行广告的投放的管理，并利用算法和技术自动实现精准的目标受众定向,只把广告投放给对的人。

对广告主而言，使用程序化购买不仅能提高流量采购效率，更低成本的或可靠，稳定的流量，还能使用各种策略，投放不同的目标人群,使得广告的投放效果更加可控。减少广告浪费的同时,提升转化率，扩大覆盖面积，最终实现跨终端的投放管理和效果追踪

对媒体资源而言，使用程序化管理的方式，能实现资源的自动化售卖，提高流量的使用率，有效利用优质流量和长尾流量，并且能基于人群的属性、兴趣等标签,对不同流量给出不同售价,提升流量库存的收入。

DSP 怎么做的

简单讲，就是DSP平台先对接媒体，如微博、腾讯、头条、OPPO、ZAKER等资源，再通过实时竞价（RTB）的方式，让广告主自主选择投放的媒体及投放的客户群体，进行广告投放~

DSP投放系统架构图

平台核心要素

1) 流量资源

不同的流量资源会有不同的效果,如何选择媒体和广告位是关键。平台上各种站点，频道，广告位，广告投放前制定投放策略,需要根据投放KPI(即广告主用于评估效果的关键指标)选择适合的媒体。

制定媒体策略首先需要分析目标受众,根据目标受众选择匹配的媒体资源。媒体策略可以从所要投放的广告位所在页面(page)、位置(spot)、流量 (inventory)、价格(price)和转化(conversion)五个角度来制定。

- 页面: 包括渠道、载体(PC/WAP/APP)、媒体分类、频道、URL、页面关键词 等。
- 位置: 广告位所在的具体位置、尺寸的大小、广告位类型、屏次(如首屏、第二屏)等,都会影响广告的有效性。
- 流量: 总流量的大小决定着能采购的流量大小。比如,每天需要投放3,000,000个展示,那么起码得先确保所选媒体的总流量大于这个量级。需求方平台通常有流量预估功能,可以为需求方的流量采购提供一定的参考价值。
- 价格: 广告主KPI一般会对单价或量级有要求,要在有限的投放预算里面满足 KPI,就得考虑价格。在竞价环境中,价格是不固定的,但是我们可以参考历史投放的平均价格。
- 转化: 除了参考流量和价格,还需要参考历史转化情况,以达到更好的媒体效果 预估。

2) 投放策略及受众定向

1. 基础设置

- 出价上限，可以按照CPM或CPC设置，广告主可以设置好能承受的价格上限,将投放成本控制在可接受范围。
- 预算控制，提前设置好预算可以避免消耗过度，预算分为每日预算、总预算。在平台上还可以设置曝光数限制和点击数量限制，相应的有每日曝光数、总曝光数或每日点击数、总点击数。
- 频次控制，频次控制是指广告主设置的同一个用户在设定时间内看到特定广告的总次数,或者在设

定时间内看到某个广告素材的总次数。一方面,在预算有限的情况下,有效的频次控制可以覆盖到更多的用户;另一方面,同段时间内对同一用户进行大量曝光或者大量重复同一广告素材的意义不大,有效的频次控制还能在一定程度上提升点击率和转化率。

- 投放时间, 设置投放日期或投放时段,也可以设置为不限制。投放日期或每日的投放时段都会对广告效果产生影响,比如周末、晚上高峰阶段。

1. 受众定向

- 地区定向, 指定投放区域,包括国家、省份、城市等。
- 客户端定向, 设置受众的操作系统、浏览器、系统语言等。
- 移动设备, 包括设备类型(如手机、平板)、设备品牌(如 Apple、华为)和设备 型号等。
- 人群标签定向, 基于DMP系统, 可以选着不同兴趣爱好的人群进行广告投放, 以实现更精准的人群覆盖

3) 创意内容

选择好广告位和对应的人群,还需要有适合的创意匹配用户的喜好。即使是同一个广告位,甚至是同一个用户,对不同创意素材的反应都是不同的。制作创意首先要把握好创意形式(Format)和用户交互(Engagement)方式。

创意形式是指素材是以什么文件形式呈现,比如图片、文字、图文、Flash、视 频、表单或者其它。目前来看,视频创意和图文创意(如信息流)创意形式最佳,这 跟广告位特性有关,这些广告位的价格也比普通广告位高一些。用户交互是指用户在素材上进行什么动作,如点击切换动态创意、在创意上填写注册信息、点击广告后扩展素材、甚至是语音提交信息。

- 图片PC
- 小灯泡
- 动态创意

数据指标

作为广告主一方, 在广告投放过程中, 必须要需要关注广告的投放效果, 以及时调整投放策略, 达到效果最大化, 关注的指标包含: 请求, 展示, 点击, 花费 基础指标, 以及 不同时间段, 不同网站, 策略, 创意, 广告形式, 省市, 频次 的高阶指标

1. 基础指标

- 请求数: 包含总请求数, 有效请求数, 参与竞价数, 赢得竞价数
- CPC: 每点击成本, 即广告被点击一次对应的价格。用公式表示即 $CPC = Cost/Click$ 。如果结算类型是CPC, 那么CPC和CTR会是主要的考核和优化目标
- CPM: 每千人成本, 即广告被展示一千次对应的价格。用公式表示即 $CPM = (Cost/Impression) * 1000$ 。如果结算类型是CPM, 一般会从CPM和CTR着手
- CTR: 指广告的点击率,即广告点击次数占广告展示次数的百分比。用公式表示即 $CTR = (Click/Impression) * 100\%$

2. 高阶指标

- 独立访客数(UV): Unique Visitor数量,指在特定时间内访问页面的虚拟自然人(用客户端标识)的总数。
- 访问数 ...
- 网站到达率: 指用户到达landingpage的数量和点击数的比值, 用于衡量该环节用户流失率

- 频次：分为离线频次和累计频次，指对每个独立用户的曝光频次
- 网站浏览量：用户到达目标页面产生的浏览量
- 二跳率：...
- 平均停留时间：该次投放用户到达目标页面后的平均停留时间
- 下载，激活，注册：该次广告投放引导用户下载，激活，注册的数量
- 付款金额，付款人次，该次广告投放引导用户产生支付的金额，人次

需求分析

1. 广告主要看实时统计广告效果数据，并且支持多维度分析
2. 把服务端广告竞价阶段的信息补充到客户端上报的日志里，需要实时日志拼接
3. 还有最基本的保证数据的准确性

架构选型

- 技术选型(1) 方案一：
 - 广告请求时把信息编码到广告url里面，曝光和点击的时候再上报。
 - 缺点：
 - 广告url过长，“偷”用户流量，移动端优质APP难以接受。
 - 不能放入过多的信息。
- 技术选型(2) 方案二：
 - 在实时处理引擎里面实现流式关联
 - 缺点：
 - join 窗口的双流数据都是被缓存在内存中的，也就是说如果某个key上的窗口数据太多就会导致 JVM OOM。
 - 广告下发后的曝光，转化等数据可能需要几天之后才能回传回来，内存无法缓存如此长时间的请求数据。
- 技术选型(3) 方案三：
 - 广告请求信息写入kv store，广告曝光和点击时查询kv store，做日志关联
 - 优点：
 - 不需要使用移动端回传
 - kv store里保存近1个月的数据，可保证拼接率

数据管理平台整体流程介绍

1. 为什么要有这个平台，平台的功能
 1. 服务于DSP系统模型
 2. 服务于广告主看数需求
2. 采集
 1. 竞价引擎日志如何采集、如何进kafka、为啥进kafka、如何保障稳定性
 2. client用户上报日志如何采集、如何进kafka、不落地如何保证数据不丢
 3. 广告主回传数据：奈学在微信朋友圈投广告，为什么要回传，如何回传
3. 处理

4. 结果展示

平台的架构难度

1. 实时拼接
 1. 拼接率
 2. 拼接失败重试方案
 3. 准确性
2. 读取kv的压力
3. 数据重复问题
4. 多维数据分析

如何从架构设计上解决这些问题

1. *实时引擎选型
2. 拼接失败重试方案
3. 实时拼接，采用离线来修复
4. 存储引擎选型，读取kv的压力，读kv慢，flink处理能力下降，数据处理不过来
5. 数据重复问题，日志加event_id
6. *多维分析引擎选型

资源评估
